RESEARCH ARTICLE

JOURNAL OF
Hymenoptera
RESEARCH
The International Society of Hymenopterists

Molecular data support *Bombus sonorus* and *Bombus pensylvanicus* (Hymenoptera, Apidae) as distinct species

Jessica L. Beckham[1], Jeff A. Johnson[2,3], Russell S. Pfau[4]

1 *Department of Integrative Biology, University of Texas at San Antonio, One UTSA Circle, San Antonio TX, 78249, USA* 2 *Department of Biological Sciences, University of North Texas, 1155 Union Circle #305220, Denton, TX 76203, USA* 3 *The Peregrine Fund, The World Center for Birds of Prey, 5668 West Flying Hawk Lane, Boise, ID 83709, USA* 4 *Department of Biological Sciences, Tarleton State University, Stephenville, TX, 76402, USA*

Corresponding author: Russell S. Pfau (rspfau@proton.me)

Citation: Beckham JL, Johnson JA, Pfau RS (2024) Molecular data support *Bombus sonorus* and *Bombus pensylvanicus* (Hymenoptera, Apidae) as distinct species. Journal of Hymenoptera Research 97: 895–914. https://doi.org/10.3897/jhr.97.132937

Abstract

Despite their distinctive banding patterns, setal coloration, and geographic ranges, the Sonoran bumble bee (*Bombus sonorus* Say, 1837) and the American bumble bee (*Bombus pensylvanicus* De Geer, 1773) are often treated as conspecific, with some authorities ranking *B. sonorus* as a subspecies of *B. pensylvanicus*. This lack of taxonomic clarity creates challenges, particularly for population monitoring and conservation initiatives. In this study, we used genetic analyses to assess the potential for clinal variation, ongoing hybridization, and historical introgression between *B. pensylvanicus* and *B. sonorus* within a broad area of sympatry across the state of Texas. Double digest restriction associated DNA sequencing (ddRADseq) was performed on 166 specimens (58 *B. sonorus* and 108 *B. pensylvanicus*), and a portion of the mitochondrial COI gene was sequenced for a subset of the specimens from Texas (46 *B. sonorus* and 32 *B. pensylvanicus*) and eight specimens of *B. sonorus* from California. An additional five sequences of *B. pensylvanicus* from Georgia and Florida were obtained from Genbank and BOLD along with one each of *B. transversalis* (Olivier, 1789), *B. mexicanus* Cresson, 1879, *B. medius* Cresson, 1863 and *B. fervidus* (Fabricius, 1798), and *B. mesomelas* Gerstäcker, 1869. For both genetic datasets (nuclear ddRADseq and mtDNA COI), individuals formed two distinct groups concordant with species identification based on setal coloration. We found no evidence supporting a clinal pattern of variation, ongoing hybridization, or historical introgression within the study area and conclude that *B. sonorus* should be recognized as a species under the Biological Species Concept.

Keywords

Bombus pensylvanicus, *Bombus sonorus*, bumble bee, COI, RADseq, sympatry

Introduction

Bombus sonorus Say, 1837, commonly referred to as the Sonoran bumble bee, has a contentious taxonomic history. First described from Mexico (Say 1837), *B. sonorus* continued to be recognized as a species by Franklin (1913), Thorp (1983), and Williams (1998), but often with some hesitation. Others have considered *B. sonorus* as conspecific with *Bombus pensylvanicus* (De Geer 1773) and often as a subspecies (*B. p. sonorus*; Handlirsch 1888; Peters 1968; Milliron 1973; Labougle 1990; Williams et al. 2014). The range of *B. sonorus* extends from California to central Texas and south to southern Mexico, overlapping with *B. pensylvanicus* within Texas and Mexico where the two are broadly sympatric. Within this region of sympatry, there have been anecdotal reports of chromatically intermediate individuals in Mexico and southwestern Texas (Labougle 1990; Williams et al. 2014). However, we are aware of only one description that has been provided as to what constitutes an intermediate phenotype for these two taxa: Labougle (1990) simply stated that setal coloration of the scutellum and punctation of the clypeus was intermediate. In a study of the hymenopteran fauna of Mexico, Peters (1968) did not mention specimens with an intermediate phenotype, but noted that both color forms occurred together; he reported collecting both on the same shrub and observed an individual of *B. sonorus* at the entrance of a nest along with 10 *B. pensylvanicus*. Because of this, he considered them to be conspecific and stated that he found it difficult to give them even subspecies designations. Based on these accounts, some authorities have been hesitant to recognize these as distinct species. More recently, Cameron et al. (2007) used four nuclear loci and one mitochondrial (mtDNA) locus (16S) to analyze a nearly complete sampling of *Bombus* (218 of 250 species). Their data showed that *B. sonorus* and *B. pensylvanicus* were polyphyletic, suggesting either introgression between the two taxa or the existence of shared ancestral variation (i.e., incomplete lineage sorting).

Some populations of *B. pensylvanicus* and *B. sonorus* are in decline (Committee on the Status of Pollinators 2007; Colla and Packer 2008; Grixti et al. 2009; Cameron et al. 2011; Colla et al. 2011, 2012; Bartomeus et al. 2013; Hatfield et al. 2015). As such, *B. pensylvanicus sensu lato* is under review for federal protection under the Endangered Species Act (USFS 2021) and is listed as Vulnerable on the IUCN Red List (Hatfield et al. 2015). Additionally, though commonly observed in Texas, both *B. sonorus* and *B. pensylvanicus* have been designated as Species of Greatest Conservation Need (SGCN) in the Texas Conservation Action Plan generated by the Texas Parks and Wildlife Department (TPWD 2012). Further study is warranted to provide additional evidence to resolve the taxonomic uncertainty of *B. sonorus* in order to help inform monitoring practices and conservation measures.

Genetic data can provide evidence to help resolve taxonomic uncertainty when morphological differences are ambiguous. Mitochondrial DNA sequences have been used to identify and delineate species among many taxonomic groups. However, the value of mtDNA sequence analysis is limited because it represents a single locus; moreover, mtDNA sequences may have a different evolutionary history relative to nuclear loci because of their maternal inheritance pattern (Moritz and Cicero 2004; Galtier et al. 2009; Taylor and Harris 2012). Other potential complications are specific to arthro-

pods, such as indirect selection on mtDNA arising from linkage disequilibrium with maternally inherited symbionts (Hurst and Jiggins 2005). An increasing number of studies, including those of *Bombus* (Duennes et al. 2012; Lozier et al. 2016; McKendrick et al. 2017), are using mtDNA sequences in combination with multi-locus nuclear genetic data. Next-generation sequencing techniques can generate large numbers of nuclear loci for the analysis of species delineation. Peterson et al. (2012) described a cost-effective reduced representation genomic sequencing approach called double digest restriction associated DNA sequencing (ddRADseq) that can produce thousands of loci from hundreds of specimens. Taken together, analyses of mtDNA sequences and ddRADseq data can provide compelling genetic insight into taxonomic relationships.

To clarify whether *B. sonorus* should be considered a distinct species, we generated ddRADseq and mtDNA cytochrome c oxidase subunit I (COI) gene sequence datasets for *B. sonorus* and *B. pensylvanicus* individuals collected across a broad area of sympatry in Texas. We sought to determine if the genetic data were concordant with species identification based on setal coloration and assessed the evidence for clinal variation, hybridization, and historical introgression.

Materials and methods

Specimen identification

For purposes of this study, we visually identified *B. sonorus* and *B. pensylvanicus* based on setal coloration (banding pattern and shade of yellow) following the descriptions of Franklin (1913), Milliron (1973), and Labougle (1990), and as depicted in Fig. 1. Specifically, the thoracic setal pile of female *B. sonorus* is black with yellow pile on the pronotum, anterior portion of scutum, and scutellum resulting in a black interalar band. In contrast, *B. pensylvanicus* usually (but not always) has a black scutellum – though some specimens bear, on the scutellum, yellow and black pile mixed, or rarely all yellow. Additionally, while T2–T3 of both taxa are yellow, T1 of *B. sonorus* is described as entirely yellow, while T1 of *B. pensylvanicus* is black with the posterior portion yellow. Lastly, *B. sonorus* exhibits a much deeper (golden) yellow than *B. pensylvanicus* (which has been described as being canary or lemon yellow).

Geographic distribution

To determine the extent of geographic overlap of *B. pensylvanicus* and *B. sonorus* within the study area (Texas), we examined photographic documentation of occurrence data on the iNaturalist platform (www.inaturalist.org). We filtered iNaturalist observations by species and location (i.e., filtered for *B. sonorus* in Texas and *B. pensylvanicus* in Texas) and then verified species identifications for all observations using the identification criteria described above. Observations were only included in our dataset if the photograph quality allowed for species identification. We then downloaded these verified observations from the Global Biodiversity Information Facility
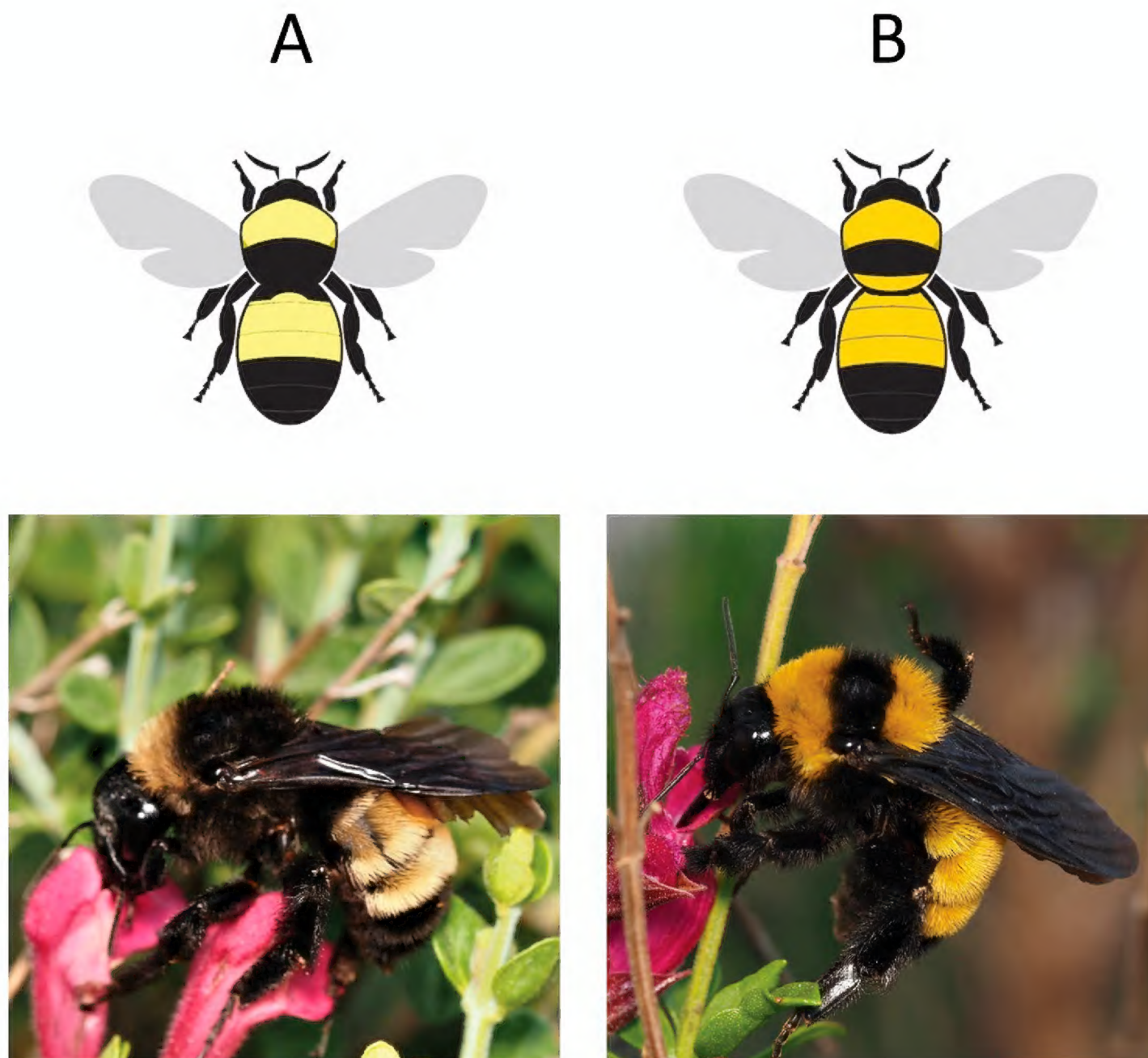
Figure 1. Coloration (banding pattern and shade of yellow) characteristic of female *B. pensylvanicus* (**A**) and *B. sonorus* (**B**) Diagram (top) modified from image provided by Texas Parks & Wildlife.

(https://www.gbif.org) and mapped each locality using QGIS (QGIS Development Team 2021). Additionally, we compared these records with the distributional map of Texas *Bombus* compiled by Warriner (2012) based on museum collections, as well as documented and projected distributions from Beckham and Atkinson (2017).

Survey area and specimen collection

To gather specimens for genetic analyses, we surveyed for bumble bees during July – September, 2016 to coincide with times of peak bumble bee activity. We identified target areas for surveys within the projected area of sympatry in Texas; these target areas were prioritized for collection trips, but we also acquired bees opportunistically from other areas of the state. Most collection events took place along public roadsides, and survey locations were chosen based on accessibility and the presence of flowering plants. Bees were collected using an aerial net and placed in a mobile cooler for transport to the lab where they were maintained at -80 °C.

We collected 166 bumble bees from 113 sites in 66 Texas counties (Fig. 2; Suppl. material 1: appendix S1); of these, all but one were female. We collected *B. sonorus* from 39 sites in 25 counties and *B. pensylvanicus* from 83 sites in 51 counties. In 11 of the 66 counties, we collected both species. Based on the distributional data (Fig. 2), 53 of the 66 counties sampled are likely within the area of sympatry in Texas (the exceptions being those counties east and south of the documented occurrence of *B. sonorus*; these include Atascosa, Bastrop, Brazos, Burleson, Caldwell, Fayette, Frio, Grayson, Lee, Milam, Montgomery, Robertson, and Wilson). None of the individuals that we collected displayed setal coloration (banding pattern and shade of yellow) intermediate between typical *B. sonorus* and *B. pensylvanicus*. Because specimens were heavily damaged during the DNA extraction process, we did not retain museum vouchers. However, photographic vouchers of most specimens can be accessed from the Dryad Digital Repository https://doi.org/10.5061/dryad.08kprr5b2. See Suppl. material 1: appendix S1 for a list of specimens examined, including locality data and mitochondrial haplotype for each.

DNA Extraction

DNA was extracted from either the head and thorax or all legs of each specimen in sterile conditions with QIAGEN® DNeasy Blood and Tissue kits using the manufacturer's protocol for DNA extraction from insect tissue, with minor adjustments. Specifically, body parts were removed from each individual bee using a sterile razor blade, and then chopped into smaller pieces. Tissue was placed in a 1.7 µl microcentrifuge tube and homogenized with a disposable micropestle prior to adding the tissue lysis buffer. Two final elutions, the first of which (Q1) was 200 µl and the second (Q2) 100 µl, were performed for each sample. DNA concentrations were determined using a Qubit™ Fluorometer. In the case of samples whose Q1 DNA concentrations were lower than 10 ng/ul, Q1 and Q2 were combined and concentrated using a Thermo Savant Speed-Vac Concentrator. Samples (Q1 or concentrated Q1 + Q2) were then prepared for sequencing in volumes that yielded approximately 2 ng of DNA.

Genomic ddRAD-sequencing

Genomic sequence data were generated using a reduced-representation genomic sequencing approach (ddRADseq) following methods described in Peterson et al. (2012). All libraries were prepared and sequenced at the Texas A&M AgriLife Genomics Facility. Briefly, genomic DNA was digested using restriction enzymes EcoRI and MspI. Barcode adaptors (110bp) were ligated to each fragment, and DNA libraries were amplified using PCR. Fragments ranging from 225–500 bases in length were size selected using a Pippen Prep instrument (Sage Science). Amplified size-selected fragment libraries were quantified and pooled in equimolar amounts, and then sequenced on two lanes of an Illumina HiSeq 4000 (150 bp paired-end reads). Two sequencing runs were performed in order to generate sufficient read depth, and raw reads of the two runs were combined after verifying that the restriction enzyme cut-sites and sample-specific barcodes were removed. The initial ddRADseq dataset (n = 166 samples) consisted of 694,934,628
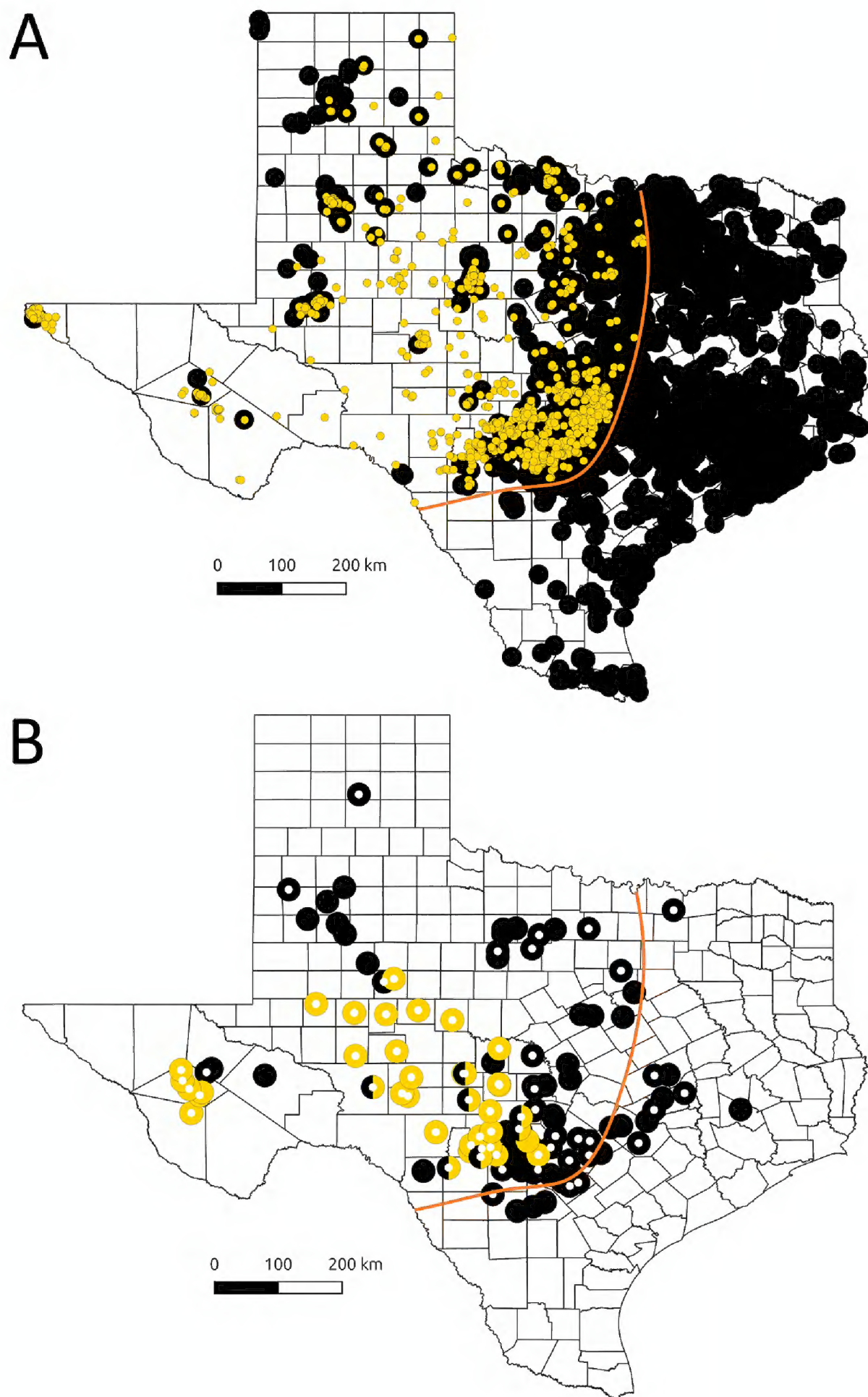
Figure 2. Occurrence of *B. pensylvanicus* (larger black circles) and *B. sonorus* (smaller yellow circles) within Texas based on iNaturalist observations (**A**) locations sampled for this study (**B**) showing locations with *B. pensylvanicus* (black circles), *B. sonorus* (yellow circles), and both species (half black and half yellow circles) as identified by ddRADseq. White dots over colored circles are locations from which mtDNA sequences were also obtained. Orange line indicates the apparent eastern- and southern-most extent of *B. sonorus* (based on iNaturalist data) for ease of comparison between the two figures.

paired demultiplexed raw sequence reads. Using Fastx Toolkit v. 0.0.13 (http://hannon-lab.cshl.edu/fastx_toolkit/), we trimmed the first five and four bases from the forward and reverse reads, respectively, to remove low quality sites based on visualization of the fastq files. We employed the ipyrad v.0.7.15 pipeline (http://github.com/dereneaton/ipyrad; Eaton 2020) for assembly, variant calling, and quality filtering.

For ipyrad, default parameters were specified unless otherwise noted for each of the seven steps used in the pipeline. Individual reads were filtered based on a minimum Phred quality score of 20 and a maximum number of Ns in a read set to 5. Reads were then de-replicated, followed by within sample de novo clustering using VSEARCH (Rognes et al. 2016). Reads were matched with a minimum sequence similarity threshold of 0.85. Clustered sequences were then aligned using MUSCLE (Edgar 2004), and rates of heterozygosity and sequencing error were then estimated across sites for subsequent consensus base calling and filtering. We assumed a maximum of 2 consensus alleles per individual and removed consensus sequences with >5 Ns per end of paired-end reads. After clustering and aligning reads for each sample to consensus sequences, the dataset was filtered based on maximum number of single nucleotide polymorphisms (SNPs) per locus (20), maximum number of indels allowed per read (8), maximum proportion of shared heterozygous sites per locus (0.5), and minimum number of samples per locus (97 for *B. pensylvanicus*, 52 for *B. sonorus*; or 90% for each group). We performed final filtering using VCFtools as follows: max-missing 0.8 (to remove loci with genotypes having ≥ 20% missing data), min-alleles 2 and max-alleles 2 (to include only biallelic sites), mac 2 (to include only sites with minor allele count ≥ 2), and thinned to retain only 1 SNP per locus.

In order to assign individuals to clusters and identify potential admixture between *B. pensylvanicus* and *B. sonorus*, we performed a principal components analysis (PCA), conducted STRUCTURE analyses (a Bayesian clustering method), and created a NeighborNet network. PCA was performed using R (R Core Team 2021) and the package adegenet 2.1.10 (Jombart 2008). The software STRUCTURE 2.3.4 (Pritchard et al. 2000; Falush et al. 2003), given a specified number of clusters (K), estimates population membership for each individual by assuming Hardy-Weinberg and linkage equilibrium within clusters. Therefore, we limited our dataset to a single SNP per RADseq locus to reduce bias due to linkage among SNPs. Each run consisted of a burn-in length of 50,000 iterations and a run length of 500,000 iterations. Each run for K = 1 to 4 was replicated ten times using an ancestry model assuming admixture, a model of correlated allele frequencies that did not include prior information on population origin, and an individual alpha for each population (Wang 2017). StrAuto (Chhatre and Emerson 2017) which implements GNU parallel (Tange 2023) was used to automate STRUCTURE runs across multiple cores of a Jetstream virtual machine (Stewart et al. 2015; Towns et al. 2014). Structure Harvester (Earl and vonHoldt 2012) was used to assess likelihood values across multiple values of K. Results of multiple STRUCTURE runs at each value of K were visualized using CLUMPAK (Kopelman et al. 2015), which implements CLUMPP (Jakobsson and Rossenberg 2007) and DISTRUCT (Rosenberg 2004) to align runs across K values and identify major and

minor modalities among runs (Jakobsson and Rosenberg 2007). We constructed the NeighborNet network using SplitsTree4 v.4.13.1 (Huson and Bryant 2006) using p-distance and 1000 bootstrap replicates.

Mitochondrial sequencing (COI)

For a subset of the Texas specimens (and eight *B. sonorus* from California), we amplified a portion of the mitochondrial cytochrome c oxidase subunit I (COI) gene using polymerase chain reaction (PCR) and sequenced (forward and reverse) the resulting PCR products using automated Sanger dideoxy chain termination sequencing. This dataset included 77 of the 166 Texas specimens (45 *B. sonorus* and 32 *B. pensylvanicus*) because COI failed to amplify or sequence in all 166 specimens, and the eight *B. sonorus* from California (Genbank accession numbers PP840056–PP840059). We carried out PCRs and sequencing reactions using the primers LCO1490 and HCO2198 (Folmer et al. 1994), targeting a region of 658 bp (excluding primers). PCR reactions occurred in a final volume of 25 μL consisting of 1X buffer, 0.16 mM deoxynucleotide triphosphates, 2.5 mM of $MgCl_2$, 0.1 μM of each primer, and 0.05 units of Taq DNA polymerase. The thermal profile consisted of a step-down procedure with the initial denaturation cycle consisting of 120 seconds at 95 °C, then 20 cycles of 30 seconds at 95 °C, 30 seconds at an initial temperature of 56 °C stepping down 0.5 °C per cycle to a 20th cycle temperature of 46 °C, then 90 seconds at 72 °C, 20 cycles of 30 seconds at 95 °C, 30 seconds at 46 °C, and 90 seconds at 72 °C followed by a final extension of 72 °C for 10 minutes.

In addition to sequences that we generated, we obtained five sequences of *B. pensylvanicus* from BOLD (http://www.barcodinglife.org; Georgia: BBBEE941-11 and BBBEE939-11; Florida: BBHYA070-12, BBHYA087-12, and BBHYA069-12) and one each of *B. transversalis* (Olivier, 1789), *B. mexicanus* Cresson, 1879, *B. medius* Cresson, 1863 and *B. fervidus* (Fabricius, 1798) from Genbank (www.ncbi.nlm.nih.gov/Genbank; KJ848950, MG279409, KC853364, and MG448456).

We aligned sequences using ClustalW within the program MEGA-X (Kumar et al. 2018) and trimmed the alignment to omit missing sites at each end resulting in 520 bp. We constructed a phylogenetic tree using MrBayes with *B. mesomelas* Gerstäcker, 1869 as an outgroup (Genbank HQ563805) based on its relationship to the other species included (Cameron et al. 2007). The best-fit substitution model was determined to be HKY in jModelTest v.2.1.10 (Guindon and Gascuel 2003; Darriba et al. 2012) using BIC. Within MrBayes, the sample and print frequency was set to 500, the diagnostic frequency was 5,000, and the run length was 1,000,000. Trees were summarized to produce posterior probabilities of each split and branch lengths. We visualized the resulting tree using FigTree 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). Uncorrected mean genetic distances (p-distances) between haplotypes were generated using the program MEGA-X (Kumar et al. 2018). We created a median-joining haplotype network using POPART 1.7 (Bandelt et al. 1999; Leigh and Bryant 2015) using uncorrected p-distance.

Results

Geographic distribution

The iNaturalist occurrence data included 2,008 *B. sonorus* (GBIF.org 2024a; retrieved on 20 April 2024 and included observations spanning the years 2005–2024) and 11,984 *B. pensylvanicus* (GBIF.org 2024b; retrieved on 20 April 2024 and included observations spanning the years 1972–2024 [only two observations predated 2004]). iNaturalist records documented the occurrence of *B. pensylvanicus* state-wide and were largely consistent with Warriner (2012), which was based on vouchered museum specimens, and Beckham and Atkinson (2017), which included both vouchered museum specimens and citizen science data. For *B. sonorus*, however, iNaturalist records showed a broader distribution than reported by both Warriner (2012) and Beckham and Atkinson (2017), extending further north and northwest than reported in either publication. Also, two records reported by Warriner (2012) appeared extralimital relative to iNaturalist records. We confirmed one of these records as being in error (the locality was misinterpreted; see Discussion).

ddRADseq analyses

We obtained an average of 4.2 million paired-end reads per sample after *q*-score and adapter filtering. Sequencing reads are available on the NCBI Short Reads Archive (BioProject ID PRJNA1115193). After de novo assembly and subsequent filtering using ipyrad, 17,352 loci were identified. The final dataset consisted of 3,371 loci after further filtering using VCFtools. We conducted all subsequent analyses using this final dataset.

The PCA showed two well-separated clusters of individuals (Fig. 3). STRUCTURE assigned individuals to two groups for K = 2, 3, and 4 (Fig. 4A; Suppl. material 2), with consistency across all 10 runs (no multimodality). At K = 2, membership coefficients were >0.842 for *B. pensylvanicus* ($\bar{x}$ = 0.989, SD = 0.026) and >0.942 for *B. sonorus* ($\bar{x}$ = 0.995, SD = 0.013). The NeighborNet network clearly separated samples into two groups (Fig. 4B). All three analyses were concordant with species identification of each individual based on coloration.

Mitochondrial (COI) analyses

Of the 90 COI sequences analyzed, we identified five COI haplotypes: one from *B. sonorus* and 4 from *B. pensylvanicus* (Fig. 5; Suppl. material 1: appendix S1; Genbank accession numbers PP840056–PP840059; BOLD accession number BBHYA069-12). Within the Bayesian phylogenetic tree and haplotype network, all specimens clustered according to species (as defined by coloration; Fig. 5). Within the haplotype network, haplotypes of the two species were separated by five mutational steps. The mean intraspecific genetic distance (p-distance) was 0.10% for *B. pensylvanicus* and 0% for *B. sonorus*. The mean interspecific genetic distance (p-distance between the two species) was 1.03%.
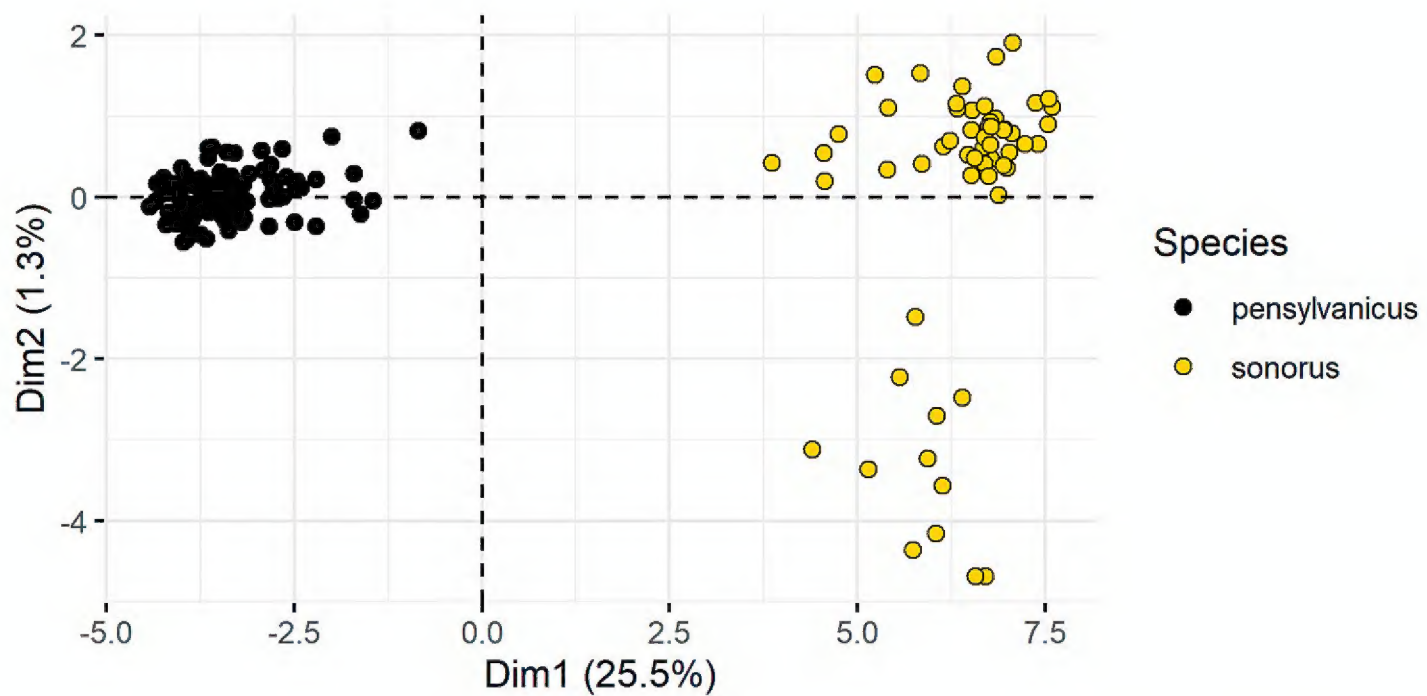
Figure 3. Principal components analysis based on ddRADseq of Texas *B. pensylvanicus* and *B. sonorus*.

Discussion

We evaluated species boundaries of the taxonomically controversial bumble bees *B. sonorus* and *B. pensylvanicus* within a broad area of sympatry using two genetic datasets (nuclear ddRADseq and mitochondrial COI) combined with examination of setal coloration (banding pattern and shade of yellow). To aid in interpreting our genetic data, we documented the extent of geographic overlap of *B. sonorus* and *B. pensylvanicus* using occurrence data from the iNaturalist platform. While the iNaturalist dataset is limited to presence data, we relied on these data because there are numerous records documented within the past two decades, each of which can be assessed visually to verify correct species identification. We found the range of *B. sonorus* documented in iNaturalist to be broader than that reported by either Warriner (2012) or Beckham and Atkinson (2017); this could be due to the different datasets used for each study or might suggest a *B. sonorus* range expansion. Warriner (2012) documented Texas bumble bee distributions primarily using historical data from vouchered museum specimens. In Warriner (2012), two county records of *B. sonorus* in eastern Texas (Hunt and Grimes counties) appeared to be extralimital. We reached out to the collection curators to verify locality data and species identification of these two records. The record reported to be *B. sonorus* from Hunt Co. had actually been collected in Hunt, TX which is in Kerr Co., Texas (well within the known distribution of *B. sonorus*). The Grimes Co. specimen of *B. sonorus* appears to be valid and was collected in 1961; however, the absence of any other specimens of *B. sonorus* this far east casts doubt on this record. Beckham and Atkinson (2017) also relied heavily on museum specimens, incorporating the data gathered by Warriner (2012) and additional museum records; a small amount of citizen science data available at the time was also included. Many museums identify *B. sonorus* as *B. pensylvanicus*, and so it is possible that both papers
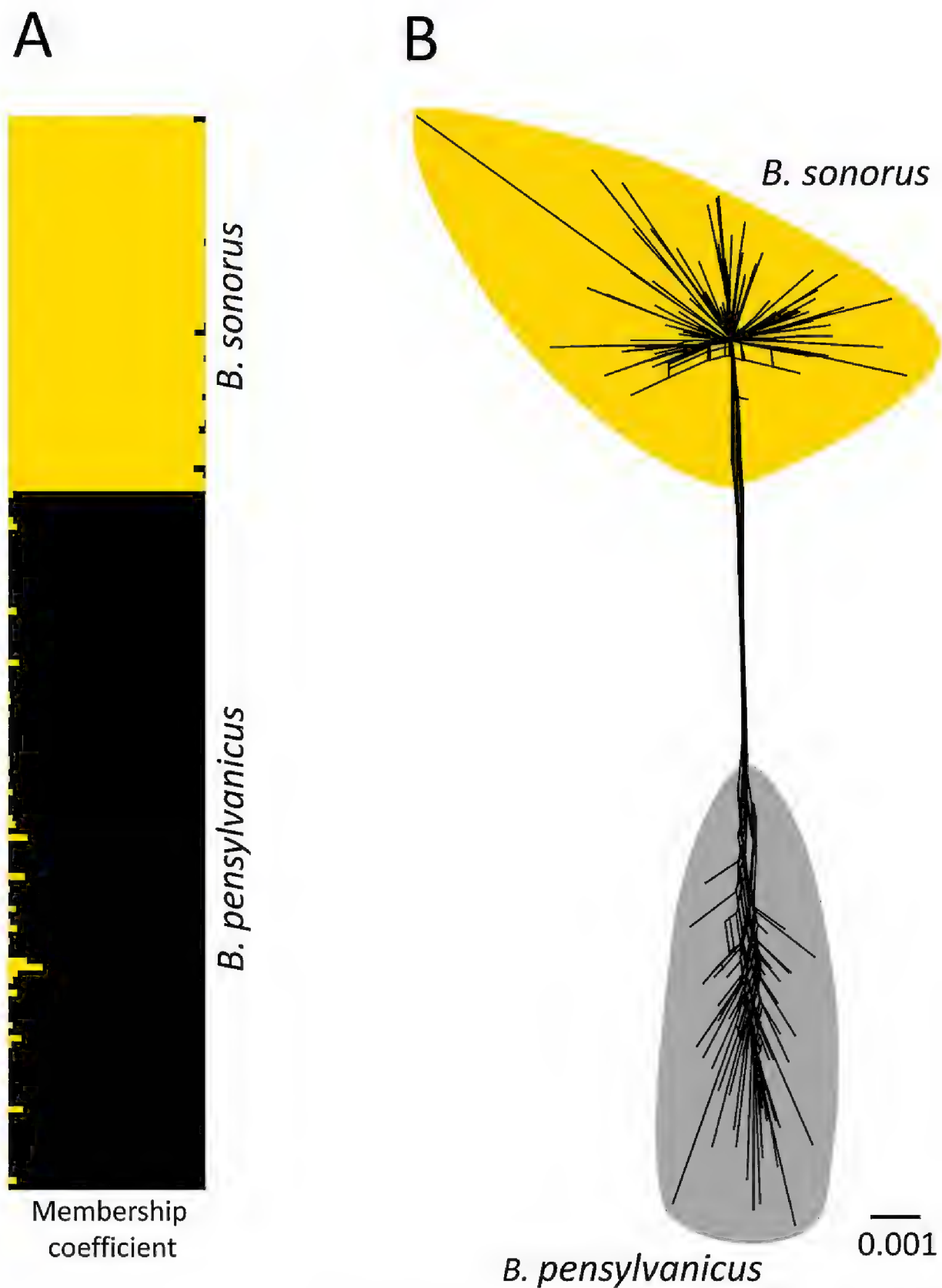
Figure 4. Results of STRUCTURE analysis (bars indicate membership coefficients) (**A**) and Neighbor-Net network (**B**) of Texas *B. pensylvanicus* and *B. sonorus* based on ddRADseq.

underreported the presence of *B. sonorus* and may have also overreported the presence of *B. pensylvanicus*. Additionally, the range of *B. sonorus* may be expanding north and east. Since the study by Beckham and Atkinson (2017), many more iNaturalist records of Texas bumble bees have been uploaded. Based on the totality of evidence, *B. sonorus* occurs in the western two-thirds of Texas (Fig. 2). In contrast, *B. pensylvanicus* occurs throughout Texas, including as far west as the vicinity of El Paso (Fig. 2). Thus, the area of sympatry includes the entirety of the distribution of *B. sonorus* in the state of Texas.
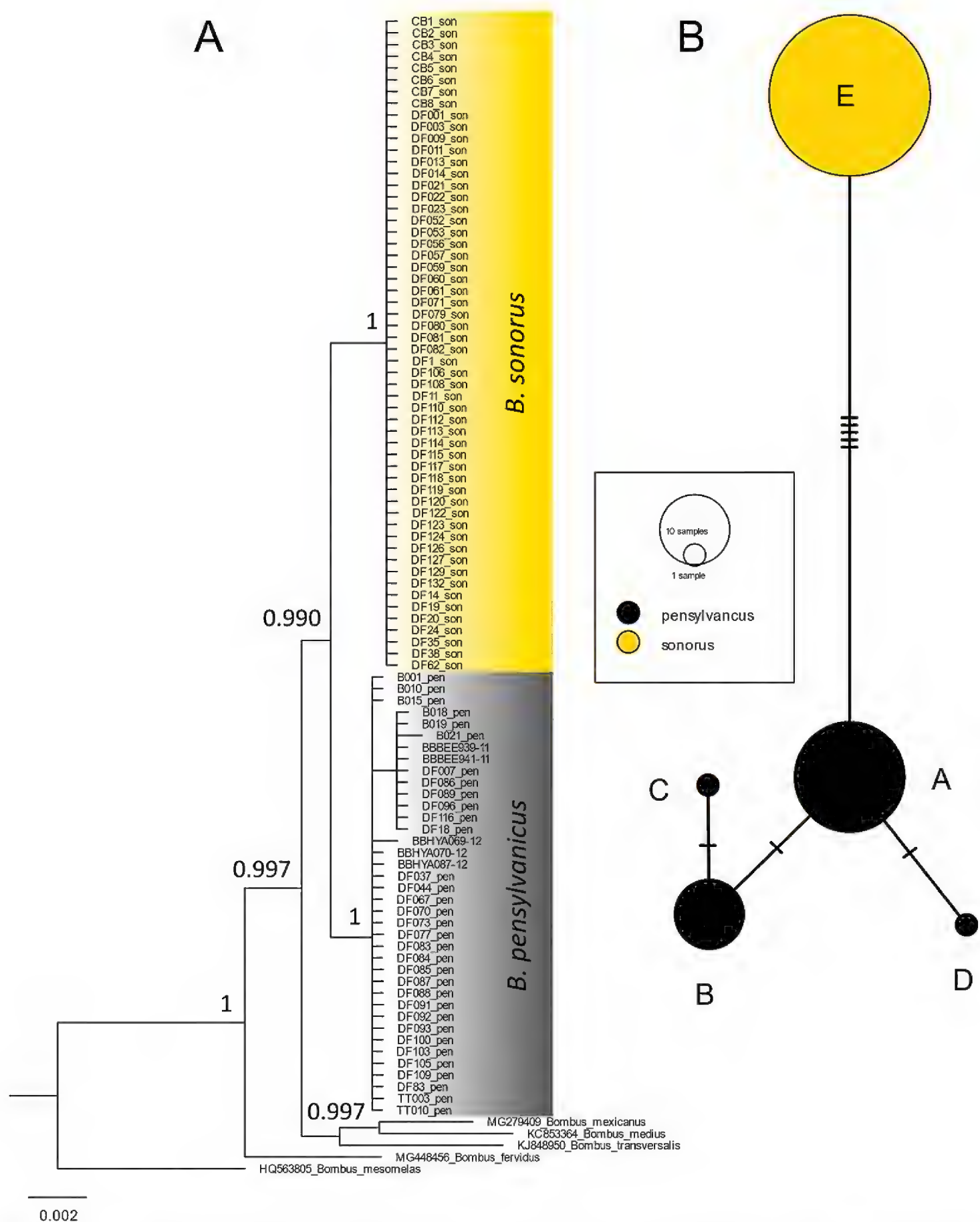
Figure 5. Bayesian phylogenetic tree of individuals (**A**) and median-joining haplotype network (**B**) of *Bombus* based on COI mtDNA sequences. The tree was rooted with *B. mesomelas*. Numbers along branches in the phylogenetic tree indicate support (posterior probability; provided only for major clades). Within the network, sizes of circles are proportional to haplotype frequency and lines across branches represent mutational steps between haplotypes. The color of circles in the network indicates species in which the haplotype was found.

All of our analyses were consistent in providing evidence that *B. sonorus* and *B. pensylvanicus* are genetically distinct across the documented area of sympatry in Texas. We found no evidence of a pattern of clinal variation, hybridization, or historical

introgression within the broad area of sympatry sampled for this project (Texas). If these color forms represent a single species, or if extensive interbreeding occurred, we would have expected one of three patterns: (1) a single genetic group, (2) individuals not grouping concordant with coloration, or (3) individuals having mixed membership to more than one cluster (for example, an F1 hybrid between two divergent lineages would have approximately 50% membership to each of two clusters in STRUCTURE). Instead, all analyses showed two distinct genetic groups. The mitochondrial COI dataset showed all individuals forming reciprocally monophyletic clades in concordance with their identification as *B. sonorus* or *B. pensylvanicus* according to coloration. Additionally, we found no instances of individuals identified as one species by ddRADseq possessing a mtDNA haplotype of the other species. If the two forms hybridize extensively and produce fertile offspring within the area of sympatry, historical introgression of mitochondria (due to backcrossing) would be expected, resulting in a lack of concordance between datasets. The lowest STRUCTURE membership coefficient was 0.842, which could represent a hybrid backcross individual; however, incomplete lineage sorting (the presence of shared variation predating speciation) results in patterns such as this which can be attributed falsely to hybridization (Hey 2006). Furthermore, perfect assignment of all individuals is never seen in this type of analysis. Given the low COI sequence divergence between *B. sonorus* and *B. pensylvanicus*, combined with no evidence of mtDNA introgression, incomplete sorting of nuclear alleles is to be expected.

Our results showed the two taxa as reciprocally monophyletic. This contrasts with Cameron et al. (2007) which showed a polyphyletic relationship between *B. sonorus* and *B. pensylvanicus*. Their study addressed the entire *Bombus* genus (thus included only two specimens each of *B. sonorus* and *B. pensylvanicus*) and used a concatenation of four nuclear genes and the mitochondrial 16S rRNA gene. Given the lower overall substitution rate of these genes, combined with the close relationship between these two species (as shown by the 1.04% COI divergence from our study), the polyphyletic pattern in Cameron et al. (2007) could be explained by incomplete lineage sorting, a well-known phenomenon among recently diverged species (Ferrer Obiol et al. 2021; DeRaad et al 2023; Nabholz 2024). We do not interpret a particular threshold of COI divergence as indicative of species status. There are numerous examples that show lack of mtDNA divergence among species that can be delineated by other means (Wiemers and Fiedler 2007; Ammerman et al. 2016; Nabholz 2024). For purposes of our study, the presence of reciprocal monophyly (regardless of degree of divergence) which is concordant with coloration and RADseq data provides evidence that hybridization (either ongoing or historical) is minimal. We interpret the low degree of divergence as evidence that these taxa are sister species that evolved from a common ancestor relatively recently.

All specimens that we collected for this study were clearly identifiable to species based on coloration, and we observed no phenotypically intermediate specimens among our samples. Milliron (1973) mentioned that the difference in shade of yellow was less in areas of sympatry. However, among the specimens we collected the shade of yellow was consistent across all individuals. Labougle (1990) reported that chromatically intermediate specimens occur, mainly in northeastern Mexico and southwestern Texas, referring to Mexican specimens as being difficult to place in either taxon because of

color of the scutellum and punctation of the clypeus intermediate between them. However, within their key, Labougle (1990) stated "scutellum and T-1 sometimes with an even mixture of black and yellow hairs". What constituted "chromatically intermediate" was not described. Among our samples, we did not observe specimens that we would describe as chromatically intermediate; the color of the scutellum was always distinct, as was the shade of yellow. Given that our sampling efforts did not extend into Mexico, we cannot be certain that the intermediates described by Labougle (1990) do not exist. There is a great need for future work in Mexico to verify the existence of intermediates.

Upon examining photographic records on iNaturalist, we did find several instances of female *B. pensylvanicus* in the U.S. with numerous yellow hairs on the scutellum (GBIF. org 2004c), as mentioned by Milliron (1973) and Labougle (1990) in their description of *B. pensylvanicus*. Importantly, these specimens were documented far from areas of sympatry between the two species (Florida, Georgia, Mississippi, and Arkansas) and, therefore, this coloration could not have been acquired by interbreeding with *B. sonorus*. We urge caution in using chromatic intermediates as evidence of hybridization given that female specimens with numerous yellow hairs on the scutellum occur far from areas of sympatry.

The observation by Peters (1968) of an individual of *B. sonorus* at the entrance of a nest along with 10 *B. pensylvanicus* was interpreted as evidence that these are color variants of the same species. However, such an interaction should not be used to determine species status. Plath (1927) found *B. terricola* Kirby, 1837 workers in 6 of 36 nests of *B. affinis* Cresson, 1863 which he explained by "natural requeening." Additionally, several instances of attempted mating involving a species of *Bombus* and another species have been photo-documented on iNaturalist (GBIF 2004d). These include *B. impatiens* Cresson 1863 with *Xylocopa virginica* (Linnaeus, 1771), *Andrena* Fabricius, 1775 with *B. griseocollis* (De-Geer, 1773), *B. fervidus* with *B. bimaculatus* Cresson, 1863, *B. impatiens* with *B. fraternus* (Smith, 1854), and *B. flavifrons* Cresson, 1863 with *B. melanopygus* Nylander, 1848. Based on these observations, it is not surprising to also see photo-documentation of a *B. sonorus* male attempting to mate with a *B. pensylvanicus* female (GBIF.org 2024e). Lastly, successful hybridization between recognized species of *Bombus* has been documented but has not been used as evidence that the species were morphological variants of the same species (Kanbe et al. 2008; Kondo et al. 2009; Yoon et al. 2009; Yoon et al. 2011; Kubo et al. 2023). In fact, natural hybridization between recognized species of animals and plants has been well documented (Arnold 1997; Mallet 2005; Taylor and Larson 2019). A critical factor in species delimitation is not whether hybridization occurs, but to what extent.

Conclusion

Our data demonstrate that hybridization events between *B. pensylvanicus* and *B. sonorus* are absent or very rare, despite ample opportunity for mating in sympatry, justifying the status of *B. sonorus* as a distinct species under the Biological Species Concept (Mayr 1963). Nuclear ddRADseq, mitochondrial COI, and coloration were concordant in distinguishing *B. sonorus* from *B. pensylvanicus*. Given the lack of evidence supporting

clinal variation, hybridization, or historical introgression between these species across the broad area of sympatry in Texas, it seems unlikely that these phenomena would be occurring in Mexico where the two species are also reportedly sympatric. However, future research should focus on the assessment of populations of *B. sonorus and B. pensylvanicus* in Mexico, to include documentation of areas in which the two species are sympatric and confirmation of the anecdotal reports of individuals displaying intermediate coloration. Until evidence of clinal variation or extensive hybridization between *B. sonorus* and *B. pensylvanicus* is documented using multilocus genetic datasets, we recommend that *B. sonorus* be recognized as a distinct species and that conservation efforts designate them as such.

Data accessibility

Raw sequencing reads (ddRADseq) are available on the NCBI Short Reads Archive (BioProject ID PRJNA1115193). COI mtDNA sequences are available at NCBI (Genbank accession numbers: PP840056–PP840059). The ddRADseq VCF, COI sequence alignment, and photographic vouchers of specimens are available from the Dryad Digital Repository https://doi.org/10.5061/dryad.08kprr5b2.

Acknowledgments

References

Ammerman LK, Lee DN, Pfau RS (2016) Patterns of genetic divergence among *Myotis californicus*, *M. ciliolabrum*, and *M. leibii* based on amplified fragment length polymorphism. Acta Chiropterologica 18: 336–346. https://doi.org/10.3161/15081109ACC2016.18.2.003

Arnold ML (1997) Natural hybridization and evolution. Oxford University Press on Demand (Oxford, UK): 1–232. https://doi.org/10.1093/oso/9780195099744.001.0001

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16: 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Bartomeus I, Ascher JS, Gibbs J, Danforth BN, Wagner DL, Hedtke SM, Winfree R (2013) Historical changes in northeastern US bee pollinators related to shared ecological traits. Proceedings of the National Academy of Sciences 110: 4656–4660. https://doi.org/10.1073/pnas.1218503110

Beckham JL, Atkinson S (2017) An updated understanding of Texas bumble bee (Hymenoptera: Apidae) species presence and potential distributions in Texas, USA. PeerJ 5: e3612. https://doi.org/10.7717/peerj.3612

Cameron SA, Hines HM, Williams PH (2007) A comprehensive phylogeny of the bumble bees (*Bombus*). Biological Journal of the Linnean Society 91: 161–188. https://doi.org/10.1111/j.1095-8312.2007.00784.x

Cameron SA, Lozier JD, Strange JP, Koch JB, Cordes N, Solter LF, Griswold TL (2011) Patterns of widespread decline in North American bumble bees. Proceedings of the National Academy of Sciences of the United States of America 108: 662–667. https://doi.org/10.1073/pnas.1014743108

Chhatre VE, Emerson KJ (2017) StrAuto: automation and parallelization of STRUCTURE analysis. BMC Bioinformatics 18: 1–5. https://doi.org/10.1186/s12859-017-1593-0

Colla SR, Gadallah F, Richardson L, Wagner D, Gall L (2012) Assessing declines of North American bumble bees (*Bombus* spp.) using museum specimens. Biodiversity & Conservation 21: 3585–3595. https://doi.org/10.1007/s10531-012-0383-2

Colla SR, Packer L (2008) Evidence for decline in eastern North American bumblebees (Hymenoptera: Apidae), with special focus on *Bombus affinis* Cresson. Biodiversity & Conservation 17: 1379–1391. https://doi.org/10.1007/s10531-008-9340-5

Colla S, Richardson L, Williams P (2011) Bumble bees of the eastern United States. US Department of Agriculture Pollinator Partnership (Washington, D.C.): 1–103.

Committee on the Status of Pollinators in North America, National Research Council (2007) Status of Pollinators in North America. National Academy Press (Washington, DC): 1–326. https://doi.org/10.17226/11761

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: More models, new heuristics and parallel computing. Nature Methods 9: 772. https://doi.org/10.1038/nmeth.2109

DeRaad DA, McCullough JM, DeCicco LH, Hime PM, Joseph L, Andersen MJ, Moyle RG (2023) Mitonuclear discordance results from incomplete lineage sorting, with no detectable evidence for gene flow, in a rapid radiation of *Todiramphus* kingfishers. Molecular Ecology 32: 4844–4862. https://doi.org/10.1111/mec.17080

Duennes MA, Lozier JD, Hines HM, Cameron SA (2012) Geographical patterns of genetic divergence in the widespread Mesoamerican bumble bee *Bombus ephippiatus* (Hymenoptera: Apidae). Molecular Phylogenetics and Evolution 64: 219–231. https://doi.org/10.1016/j.ympev.2012.03.018

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4: 359–361. https://doi.org/10.1007/s12686-011-9548-7

Eaton DA, Overcast I (2020) ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics 36: 2592–2594. https://doi.org/10.1093/bioinformatics/btz966

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792–1797. https://doi.org/10.1093/nar/gkh340

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567–1587. https://doi.org/10.1093/genetics/164.4.1567

Ferrer Obiol J, James HF, Chesser RT, Bretagnolle V, González-Solís J, Rozas J, Riutort M, Welch AJ (2021) Integrating sequence capture and restriction site-associated DNA sequencing to resolve recent radiations of pelagic seabirds. Systematic Biology 70: 976–996. https://doi.org/10.1093/sysbio/syaa101

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology Biotech: 294–299.

Franklin HJ (1913) The Bombidae of the new world. Transactions of the American Entomological Society 38: 177–486. https://doi.org/10.5962/bhl.title.122585

Galtier N, Nabholz B, Glémin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Molecular Ecology 18: 4541–4550. https://doi.org/10.1111/j.1365-294x.2009.04380.x

GBIF.org (2024a) GBIF Occurrence Download. https://doi.org/10.15468/dl.g3k63f

GBIF.org (2024b) GBIF Occurrence Download. https://doi.org/10.15468/dl.gmfudu

GBIF.org (2024c) GBIF Occurrence Download. https://doi.org/10.15468/dl.fwppxw

GBIF.org (2024d) GBIF Occurrence Download. https://doi.org/10.15468/dl.sng27x

GBIF.org (2024e) GBIF Occurrence Download. https://doi.org/10.15468/dl.u8u39w

Grixti JC, Wong LT, Cameron SA, Favret C (2009) Decline of bumble bees (*Bombus*) in the North American Midwest. Biological Conservation 142: 75–84. https://doi.org/10.1016/j.biocon.2008.09.027

Guindon S, Gascuel O (2003) A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Systematic Biology 52: 696–704. https://doi.org/10.1080/10635150390235520

Handlirsch A (1888) Die Hummelsammlung des k. k. naturhistorischen Hofmuseums. Annalen des Naturhistorischen Museums in Wien 3: 209–250.

Hatfield R, Jepsen S, Thorp R, Richardson L, Colla S, Foltz Jordan S (2015) *Bombus pensylvanicus*. The IUCN Red List of Threatened Species 2015: e.T21215172A21215281.

Hurst GDD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proceedings of the Royal Society B: Biological Sciences 272: 1525–1534. https://doi.org/10.1098/rspb.2005.3056

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23: 254–267. https://doi.org/10.1093/molbev/msj030

Jakobsson M, Rosenberg NA (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23: 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kanbe Y, Okada I, Yoneda M, Goka K, Tsuchida K (2008) Interspecific mating of the introduced bumblebee *Bombus terrestris* and the native Japanese bumblebee *Bombus hypocrita sapporoensis* results in inviable hybrids. Naturwissenschaften 95: 1003–1008. https://doi.org/10.1007/s00114-008-0415-7

Kondo NI, Yamanaka D, Kanbe Y, Kunitake YK, Yoneda M, Tsuchida K, Goka K (2009) Reproductive disturbance of Japanese bumblebees by the introduced European bumblebee *Bombus terrestris*. Naturwissenschaften 96: 467–475. https://doi.org/10.1007/s00114-008-0495-4

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) CLUMPAK: A program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources 15: 1179–1191. https://doi.org/10.1111/1755-0998.12387

Kubo R, Asanuma Y, Fujimoto E, Okuyama H, Ono M, Takahashi J-I (2023) Cross-mating between the alien bumblebee *Bombus terrestris* and two native Japanese bumblebees, *B. hypocrita sapporoensis* and *B. cryptarum florilegus*, in the Nemuro Peninsula, Japan. Scientific Reports 13: 11506. https://doi.org/10.1038/s41598-023-38631-7

Kumar S, Stecher G, Li M, Khyaz C, Tamura K (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution 35: 1547–1549. https://doi.org/10.1093/molbev/msy096

Labougle JM (1990) *Bombus* of Mexico and Central America (Hymenoptera, Apidae). University of Kansas Science Bulletin 54: 35–73.

Leigh WL, Bryant D (2015) POPART: Full-feature software for haplotype network construction. Methods in Ecology and Evolution 6: 1110–1116. https://doi.org/10.1111/2041-210X.12410

Lozier JD, Jackson JM, Dillon ME, Strange JP (2016) Population genomics of divergence among extreme and intermediate color forms in a polymorphic insect. Ecology and Evolution 6: 1075–1091. https://doi.org/10.1002/ece3.1928

Mallet J (2005) Hybridization as an invasion of the genome. Trends in Ecology & Evolution 20: 229–237. https://doi.org/10.1016/j.tree.2005.02.010

McKendrick L, Provan J, Fitzpatrick Ú, Brown MJF, Murray TE, Stolle E, Paxton RJ (2017) Microsatellite analysis supports the existence of three cryptic species within the bumble bee *Bombus lucorum sensu lato*. Conservation Genetics 18: 573–584. https://doi.org/10.1007/s10592-017-0965-3

Milliron HE (1973) A monograph of the Western Hemisphere bumblebees (Hymenoptera: Apidae; Bombinae). II. The genus *Megabombus* subgenus *Megabombus*. Memoirs of the Entomological Society of Canada 89: 81–237. https://doi.org/10.4039/entm10589fv

Moritz C, Cicero C (2004) DNA Barcoding: Promise and Pitfalls. PLoS Biology 2: e354. https://doi.org/10.1371/journal.pbio.0020354

Nabholz B (2024) Incomplete lineage sorting explains the low performance of DNA barcoding in a radiation of four species of Western European grasshoppers (Orthoptera: Acrididae: Chorthippus). Biological Journal of the Linnean Society 141: 33–50. https://doi.org/10.1093/biolinnean/blad106

Peters DS (1968) Beitrage zur Kenntnis aculeater Hymenopteren von Mexiko. I. Apinae (Apidae, Apoidea). Senckenbergiana Biologica 49: 237–248.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An inexpensive method for de dovo SNP discovery and genotyping in model and non-model species. PLoS ONE 7: e37135. https://doi.org/10.1371/journal.pone.0037135

Plath OE (1927) Notes on the nesting habits of some of the less common new England bumblebees. Psyche: A Journal of Entomology 34: 122–128. https://doi.org/10.1155/1927/43469

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959. https://doi.org/10.1093/genetics/155.2.945

QGIS Development Team (2021) QGIS Geographic Information System. Open Source Geospatial Foundation. http://qgis.org

R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ 4: e2584. https://doi.org/10.7717/peerj.2584

Stewart CA, Cockerill TM, Foster I, Hancock D, Merchant N, Skidmore E, Stanzione D, Taylor J, Tuecke S, Turner G, Vaughn M, Gaffney NI (2015) Jetstream: a self-provisioned, scalable science and engineering cloud environment. In Proceedings of the 2015 XSEDE Conference Scientific Advancements Enabled by Enhanced Cyberinfrastructure, St. Louis, Missouri (USA), July 2015. Association for Computing Machinery: 1–8. https://doi.org/10.1145/2792745.2792774

Tange O (2023) GNU Parallel 20230122 ('Bolsonaristas'). https://doi.org/10.5281/zenodo.7558957

Taylor HR, Harris WE (2012) An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. Molecular Ecology Resources 12: 377–388. https://doi.org/10.1111/j.1755-0998.2012.03119.x

Taylor SA, Larson EL (2019) Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. Nature Ecology & Evolution 3: 170–177. https://doi.org/10.1038/s41559-018-0777-y

TPWD [Texas Parks and Wildlife Department] (2012) Texas Conservation Action Plan 2012–2016: Statewide/Multi-region Handbook. In: Connaly W (Ed.) TPWD (Austin, TX, USA): 1–78.

Thorp RW, Horning Jr. DS, Dunning LL (1983) Bumble bees and cuckoo bumble bees of California (Hymenoptera, Apidae). University of California Press (Los Angeles): 1–79.

Towns J, Cockerill T, Dahan M, Foster I, Gaither K, Grimshaw A, Hazlewood V, Lathrop S, Lifka D, Peterson GD, Roskies R, Scott JR, Wilkins-Diehr N (2014) XSEDE: Accelerating scientific discovery. Computing in Science & Engineering 16: 62–74. https://doi.org/10.1109/MCSE.2014.80

Wang J (2017) The computer program STRUCTURE for assigning individuals to populations: easy to use but easier to misuse. Molecular Ecology Resources 17: 981–990. https://doi.org/10.1111/1755-0998.12650

Warriner MD (2012) Bumble bees (Hymenoptera: Apidae) of Texas: historical distributions. The Southwestern Naturalist 57: 442–445. https://doi.org/10.1894/0038-4909-57.4.442

Wiemers M, Fiedler K (2007) Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae). Frontiers in Zoology: 1–16. https://doi.org/10.1186/1742-9994-4-8

Williams PH (1998) An annotated checklist of bumble bees with an analysis of patterns of description (Hymenoptera: Apidae, Bombini). Bulletin-Natural History Museum Entomology Series 67: 79–152.

Williams P, Thorp RW, Richardson L, Colla SR (2014) Bumble bees of North America: an identification guide. Princeton University Press (Princeton, NJ): 1–208.

Yoon HJ, Kim SY, Lee KY, Lee SB, Park IG, Kim IS (2009) Interspecific hybridization of the bumblebees *Bombus ignitus* and *B. terrestris*. International Journal of Industrial Entomology 18: 41–48.

Yoon H-J, Park I-G, Lee K-Y, Kim M-A, Jin B-R (2011) Interspecific hybridization of the Korean native bumblebee *Bombus hypocrita sapporoensis* and the European bumblebee *B. terrestris*. International Journal of Industrial Entomology 23: 167–174. https://doi.org/10.7852/ijie.2011.23.1.167

Supplementary material 1

Specimens of *Bombus* examined for this study including locality data, species, sex, whether RADseq data was collected, and mtDNA haplotype

Authors: Jessica L. Beckham, Jeff A. Johnson, Russell S. Pfau

Data type: xlsx

Link: https://doi.org/10.3897/jhr.97.132937.suppl1

Supplementary material 2

Clumpak output from STRUCTURE run files of RADseq data

Authors: Jessica L. Beckham, Jeff A. Johnson, Russell S. Pfau

Data type: pdf

Link: https://doi.org/10.3897/jhr.97.132937.suppl2